

Brügelmann, Hans

## Individualisierung durch Standardisierung? Wege aus den pädagogischen Widersprüchen vermessener Schulen

*formal überarbeitete Version der Originalveröffentlichung in:*

*formally revised edition of the original source in:*

*Müller, Gundi [Hrsg.]; Sloot, Annegret [Hrsg.]: Aufklärung, Mut und Leidenschaft! Plädoyer für eine demokratische Schule. Dokumentation 74. Pädagogische Woche 5. bis 9. November 2018, Cuxhaven-Duhnen. GEW Lüneburg 2019, S. 55-74*



Bitte verwenden Sie in der Quellenangabe folgende URN oder DOI /

Please use the following URN or DOI for reference:

urn:nbn:de:0111-pedocs-168163

10.25656/01:16816

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-168163>

<https://doi.org/10.25656/01:16816>

### Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

### Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

### Kontakt / Contact:

**peDOCS**  
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation  
Informationszentrum (IZ) Bildung  
E-Mail: [pedocs@dipf.de](mailto:pedocs@dipf.de)  
Internet: [www.pedocs.de](http://www.pedocs.de)

Digitalisiert

Mitglied der

  
Leibniz-Gemeinschaft

## Individualisierung durch Standardisierung? Wege aus den pädagogischen Widersprüchen vermessener Schulen<sup>1</sup>

Dr. Hans Brügelmann

In der ZEIT Nr. 40 v. 27.9.2018 beklagt die Bildungsforscherin NINA KOLLECK, „Bildungspolitiker ignorier[t]en die Erkenntnisse der Wissenschaft“. Meine Gegenthese: Bildungspolitik ist gut beraten, Forschungsergebnisse in Ruhe zu prüfen und sich medial hoch gespielten „Folgerungen“ nicht vorschnell zu unterwerfen. Ich will die von KOLLECK selbst genannten Schwierigkeiten, die Befunde empirischer Forschung in der Bildungspolitik zu nutzen, an ihrem Beispiel Rechtschreibung illustrieren.



### Forschungsbefunde sind hoch interpretationsbedürftig

Vor einigen Jahren wurden Befunde aus einer Studie des Siegener Germanisten WOLFGANG STEINIG veröffentlicht, die dafür sprachen, dass die Rechtschreibleistungen von Grundschulkindern von 1972 bis 2002 und dann noch einmal bis 2012 deutlich abgenommen hätten (Steinig u. a. 2009; Steinig/Betzel 2014). Diese Ergebnisse erreichten eine große öffentliche Aufmerksamkeit – vom SPIEGEL mit dem Etikett „Rechtschreipkaterstrofe“ dramatisch auf den Punkt gebracht. Einbezogen in die Untersuchung waren aber jeweils nur zehn Klassen, zudem beschränkt auf das Ruhrgebiet. Leistungsverluste in der „unteren“ Sozialschicht wurden neuen Unterrichtsmethoden angelastet – allerdings ohne den sozio-ökonomischen Strukturwandel zu berücksichtigen: fast Vollbeschäftigung Anfang der 1970er Jahre, eine Arbeitslosigkeit<sup>1</sup> und Hartz-IV-Abhängigkeit von 15-30% der Bevölkerung nach

<sup>1</sup> Der Vortrag orientierte sich an bereits veröffentlichten Texten. So deckt sich der erste Teil dieses Beitrags weitgehend mit dem Artikel „Großes Desinteresse“ – oder kluge Zurückhaltung“ (erschieden in: Lehren & Lernen, H. 11/ 2018), im zweiten habe ich Stücke aus „Evidenzbasierung der Pädagogik. Scheinlösung für ein schwieriges Problem“ (erschieden in: Pädagogik H. 10/2015, S 38-43) aufgenommen und im Schlussteil den Hauptteil aus „Kinderrechte und Schulqualität: Plädoyer für eine pädagogische Lern- und Leistungskultur“ (erschieden in: Krappmann, L./ Petry C. (Hrsg.) (2016): Worauf Kinder und Jugendliche ein Recht haben. Kinderrechte, Demokratie und Schule: Ein Manifest. Debus Verlag: Schwalbach, S. 136-148).



der Jahrtausendwende. Von der Bildungspolitik wurde dennoch umgehend verlangt, bestimmte, in der Studie übrigens gar nicht überprüfte, Methoden wie ein „Schreiben nach Gehör“ oder „Lesen durch Schreiben“ zu verbieten.

Inzwischen veröffentlichte die Schriftsprachdidaktikerin PETRA HÜTTIS-GRAFF die Ergebnisse eines Vergleichs von Hamburger Grundschulkindern 1994 vs. 2014 (Hüttis-Graff 2015; Hüttis-Graff/ Wirsching 2018). Sie stellte in den untersuchten 18 Klassen eine Verbesserung der Rechtschreibleistungen<sup>8</sup> auf allen Klassenstufen fest. Diese forschungsmethodisch zu STEINIG gleichwertige Positivnachricht fand keine öffentliche Beachtung, nicht einmal die auffällige Differenz der Ergebnisse zur Siegener Studie wurde diskutiert: Handelte es sich um regional bedingte Unterschiede? Haben die Test-Aufgaben unterschiedliche Teilkompetenzen abgeprüft? War in den Klassen jeweils unterschiedlich unterrichtet worden?

Wie soll Bildungspolitik mit einer solchen Situation umgehen? Wir brauchen mehr große Repräsentativstudien, fordert KOLLECK. Aber auch deren Ergebnisse sind nicht leicht auf einen Nenner zu bringen. Zwar findet der IQB-Bildungstrend für die Jahre 2011 bis 2016 in allen Bundesländern einen Abfall der Rechtschreibleistungen (Stanat u. a. 2017). Aber wie passt dieser Befund zu dem bundesweiten Leistungsanstieg in der Rechtschreibung, der bei der IGLU-E-Studie für die Zeit von 2001 bis 2005 (Valtin/ Voss 2010) und für die Standardisierung der Hamburger Schreibprobe von 2002 bis 2012 (May 2013) festgestellt worden ist? Existiert tatsächlich ein Knick um 2011/12 – und vor allem wie lässt er sich ggf. erklären? „Lesen durch Schreiben“ jedenfalls hatte seine Hochphase vor der Jahrtausendwende. Ist er also zurückzuführen auf eine veränderte Zusammensetzung der Schülerschaft, andere Testaufgaben oder...? Und wie plausibel ist ein Knick überhaupt angesichts der Konstanz der Rechtschreibleistungen in der I.e.o.-Studie bei Erwachsenen (Grotluschen/Riekmann 2012), die zwischen den 1950er und den 1990er Jahren in die Grundschule gegangen sind, also in den Hochzeiten von Fibellehrgang einerseits und freiem Schreiben andererseits?

So bleibt nur die Hoffnung auf kontrollierte Vergleiche verschiedener Methoden. Ich selbst habe zu Beginn meiner wissenschaftlichen Beschäftigung mit konkurrierenden Konzepten des Anfangsunterrichts für solche Studien plädiert (Brügelmann/Söhnen 1981), bin aber inzwischen skeptisch. Aber auch die Untersuchungen zu den Wirkungen einzelner Methoden bringen keine Klarheit. Einerseits belegen Forschungsüberblicke in den USA, dass „invented spelling“, also ein lautorientiertes Verschriften von Wörtern, Kinder hilft, das alphabetische Prinzip unserer Schrift zu begreifen (National Reading Panel 2000). Außerdem korreliert

– auch in deutschen Untersuchungen (Richter 1992; Brügelmann 1999; Scheerer-Neumann Schnitzler 2009) – ein hohes Niveau des lautorientierten Schreibens am Anfang hoch mit guten Rechtschreibleistungen später (National Early Literacy Panel 2008; weitere pädagogische Argumente, empirische Befunde und didaktisch-methodische Ideen zum freien Schreiben und einem Einstieg(!) in den Schriftspracherwerb über die lautorientierte Verschriftung von Wörtern, findet sich in Brinkmann 2015 und 2018 ODER sowie in ihrer Zusammenfassung des Spracherfahrungsansatzes im Anhang zu diesem Beitrag.

Aber zeigen nicht Untersuchungen zu „Lesen durch Schreiben“, dass Kinder mit dieser Methode schlechtere Rechtschreibleistungen erbringen? Die Ergebnisse der bisherigen Methodenvergleiche fallen widersprüchlich aus (Brügelmann/Brinkmann 2006/2013), auch die einzige Metaanalyse konnte keine belastbaren Unterschiede finden (Funke 2014).

Da scheint eine in diesen Tagen viel und auch von KOLLECK zitierte Bonner Studie (idw 2018) endlich Klarheit zu schaffen, berichtet sie doch massive Fehlerunterschiede bei den verschiedenen Methoden – und das bei insgesamt 3.000 untersuchten Kindern. Die brandenburgische Kultusministerin hat deshalb nicht lange gefackelt und zwei Wochen nach der ersten Pressemitteilung einen Unterricht nach Lesen durch Schreiben verboten. Ein Vorbild ganz im Sinne von KOLLECKS Forderung, Bildungspolitik müsse sich stärker an wissenschaftlichen Erkenntnissen orientieren?

Schaut man genauer hin, wird auch hier deutlich, dass es nicht so einfach ist mit den Folgerungen aus empirischen Studien, zumal wenn ihre Anlage und Methodik so schwach dokumentiert ist wie bisher in diesem Fall (Kuhl/Röhr-Sendlmeier 2018). Im echten Längsschnitt der Bonner waren nur knapp 300 der 3.000 Kinder einbezogen, also sechs Klassen pro Methode. Da können schon ein, zwei Lehrer\*innen oder die besondere Zusammensetzung einzelner Klassen das Ergebnis stark beeinflussen. Zwar wurde angeblich der Einfluss der schriftsprachlichen Voraussetzungen der Kinder statistisch kontrolliert, aber das eingesetzte Verfahren hat nur eine geringe Prognosekraft für die spätere Rechtschreibleistung, sichert also nicht die Vergleichbarkeit der Voraussetzungen. Und wie die verglichenen Ansätze von den Lehrer\*innen tatsächlich umgesetzt wurden, haben die Forscher\*innen gar nicht erhoben, so dass es schwer ist den Geltungsbereich der Ergebnisse einzuschätzen.

Zahlen sprechen eben nicht für sich. Man muss sie interpretieren und auch gewichten, welche der beobachteten Effekte einem wichtiger sind. Vor allem reicht es nicht, Abstracts und Folgerungen der Autor\*innen zu lesen, man muss schon die oft divergenten Detailbefunde zur Kenntnis nehmen und auch die statistischen Kennwerte auf ihre Geltungskraft



überprüfen. Zum Beispiel haben in der auch von KOLLECK zitierten Berliner Studie „BeLesen“ Kinder mit Migrationshintergrund, die nach „Lesen durch Schreiben“ unterrichtet worden waren, später in der Rechtschreibung schlechter abgeschnitten als in den Fibelklassen, dafür aber im Lesen besser als nach einem Fibelunterricht (Schröder-Lenzen/Merkens 2006).

Auch in der Hamburger PLUS-Studie hatten die Fibellehrgänge auf den ersten Blick Vorteile gegenüber alternativen Methoden. Wurde aber der soziale Hintergrund kontrolliert, verschwanden die Differenzen in den Leistungen der Klassen, die nach verschiedenen Methoden unterrichtet worden waren (May 1995). Dass gerade Lehrer\*innen in eher schwierigen Stadtteilen mit offeneren Ansätzen arbeiten, zeigte sich auch in anderen Studien (Kirschhock 2004). Ein weiterer Hinweis auf die Kontextabhängigkeit von Wirkungen im pädagogischen Feld, auf die KOLLECK zu Recht hinweist und die Allgemeinaussagen und damit auch bildungspolitische Folgerungen erschweren. Ob Fallstudien oder Repräsentativstudien – sie sind wichtig, um auf Potenziale und Risiken von Interventionen aufmerksam zu machen, aber sie können eben auch nur Wahrscheinlichkeiten benennen, die je nach Umfeld unterschiedlich ausfallen werden (Brügelmann 2015).

Bedenkt man zusätzlich, dass in der Regel selbst für ein-und-dieselbe Methode die Ergebnisse verschiedener Klassen breit streuen, weil die Lehrer\*innen sie unterschiedlich umsetzen, wird noch verständlicher, warum Politik und Verwaltung mit dem Vorschreiben oder Verboten von Methoden sehr vorsichtig sein sollten. Unterschiedliche Bedingungen erfordern Flexibilität und genau darum gibt es eine Methodenfreiheit. Dass es dabei auch zu Fehlern kommt, dass nicht alle Lehrer\*innen gleichermaßen kompetent, verantwortungsvoll und engagiert sind, muss Anlass zu mehr Fortbildung und mehr Beratung durch die Schulaufsicht sein. Vor allem aber muss der Anfangsunterricht (auch in Mathematik, wo es ebenfalls zu viele Schüler\*innen mit elementaren Problemen gibt – und auch 2016 mehr gab als 2011, Stanat u. a. 2017) wieder Schwerpunkt der Ausbildung werden. Eine solche Unterstützung verspricht mehr Wirkung als Methodenverbote.

### **Forschung als kritisches Gegenüber – nicht als Autorität**

Ganz grundsätzlich gilt, dass Bildungsforschung und -politik (Baumert 2016) und Wissenschaft und pädagogische Praxis (Brügelmann 2015) je eigenen Handlungslogiken folgen. Insofern – um ein anderes Beispiel von KOLLECK aufzugreifen – macht es auch Sinn, trotz der vorliegenden Befunde nicht generell am 8jährigen Abitur festzuhalten, sondern etwa für die zweite Säule des Systems (Oberschule, Stadtteilschule o. ä.) neben dem achtjährigen Gymnasium eine neunjährige Form vorzuhalten,

um für diese Schulform eine stärkere soziale und Leistungsmischung zu erreichen. Politik muss eben nicht nur die Vielfalt der Kontexte bedenken, sondern auch Zielkonkurrenzen abwägen, die in der Forschung in der Regel ausgeblendet werden.

Damit sind wir bei der Grundfrage, wie das Verhältnis von Erziehungswissenschaft zu Bildungspolitik einerseits, zu pädagogischer Praxis andererseits sinnvoll zu bestimmen ist. Unter dem Anspruch einer „Evidenzbasierung“ wird gefordert, die pädagogische Praxis stärker an der Wissenschaft auszurichten, die sich ihrerseits orientiert am „Goldstandard“ empirischer Forschung, d. h. Studien über ein Kontrollgruppen-Design mit – oft großen – Zufallsstichproben, aus denen mit Hilfe standardisierter Messverfahren generalisierbare Daten gewonnen werden sollen. Auf diese Weise lassen sich in der Tat interessante Einblicke in pädagogische Sachverhalte gewinnen (vgl. etwa zu PISA Klieme u. a. 2010). Problematisch ist aber der Anspruch einer grundsätzlichen Überlegenheit dieses Erkenntniszugangs gegenüber anderen. Vor allem stellt sich die Frage, ob auf diese Weise Einsichten gewonnen werden können, die für praktische Entscheidungen hilfreich sind, also das Handeln in konkreten Situationen anleiten können.

Die Grundfrage: Werden Schulen besser, wenn Schulreform als Transfer von Forschung konzipiert wird, indem Wissenschaftler\*innen die Ergebnisse ihrer empirischen Studien in Handlungsanweisungen für Politik, Verwaltung und Praxis übersetzen, die diese dann möglichst vorgabengetreu („evidenzbasiert“) anwenden. Unter dem Etikett RDD (Research, Development, Diffusion), also von der (möglichst experimentellen) Forschung über die Entwicklung von Modellen und ihre Erprobung in Feldversuchen hin zu deren Verbreitung und Anwendung im Alltag, ist dieser „centre-periphery“-Ansatz bereits Anfang der 1970er Jahre kontrovers diskutiert worden (vgl. zur aktuellen Debatte vor allem Herzog 2013 und Dammer 2015). Auch heute stellt sich die Frage: Welche Probleme der pädagogischen Praxis sind in den letzten 10, 15 Jahren dank empirischer Befunde der Bildungsforschung gelöst worden, welche Schwierigkeiten können Lehrer\*innen im Alltag besser bewältigen?

Insbesondere die großen Testprogramme wie PISA und IGLU, aber auch die landesweiten Vergleichsarbeiten (VerA) sind mit dem Hinweis kritisiert worden, vom vielen Wiegen werde die Sau nicht fett, d. h. die Leistungen der Schule bzw. ihrer Schüler\*innen nicht besser. Diese Kritik trifft zu, greift aber zu kurz. Sie stellt zum ersten nicht in Frage, ob die Waage (sprich: der jeweils eingesetzte Leistungstest) überhaupt misst, was gemessen werden soll, und sie akzeptiert zum zweiten, dass eine



messbare Gewichtszunahme (sprich: Verbesserung der Testwerte) ein angemessener Maßstab für die angestrebten Lernerfolge sei.

Beide Annahmen sind problematisch – sei es auf der Ebene des Bildungssystems, sei es auf der Ebene der Schule oder auf der Ebene der einzelnen Schüler\*innen. Nicht, weil Leistungstests nutzlos wären. Im Gegenteil: Sie sind hilfreiche Instrumente, solange man ihren Status nicht überschätzt und den Geltungsbereich ihrer Ergebnisse beachtet. Überzogen wird ihr Kredit, wenn man ihre Befunde im Vergleich zur Alltagserfahrung und zum professionellen Urteil als höherwertig betrachtet – und nicht als komplementär, d. h. als hilfreiche Ergänzung.

Mit dem Anspruch einer stärkeren „Evidenzbasierung“ ihrer Praxis ist die Pädagogik in den letzten zehn Jahren unter Druck geraten (vgl. die Beiträge zu Böttcher u. a. 2009). Die Forderung: In der Schule dürften nur Methoden eingesetzt werden, die sich in empirischen Untersuchungen als „wirksam“ erwiesen haben. Der bereits erwähnte „Goldstandard“ für solche Studien verlangt nicht nur eine bestimmte Anlage der Untersuchung (experimentelles Design oder zumindest Zufallsstichproben mit Vergleichsgruppen), er setzt auch auf bestimmte Verfahren der Erhebung und Auswertung von Daten. Dabei stellen sich aber zwei Probleme (ausführlicher Brügelmann 2015):

#### **Das Durchschnittsproblem:**

##### **Menschliches Verhalten ist kontextabhängig**

Die Reichweite von Befunden der Bildungsforschung ist nicht nur zeitgebunden, sondern auch situationsabhängig. Für unterschiedliche nationale Kulturen mag dies unmittelbar einleuchten. Kontextabhängigkeit gilt aber auch für Subkulturen innerhalb einer Gesellschaft wie für das soziale Milieu der Familie oder einzelner Klassenzimmer. Was Lehrerin A gelingt, mag bei Lehrer B schiefgehen – selbst bei gleicher Kompetenz. Trotz *Super Nanny* und ihrer „Erfolge“: Unterrichtsmethoden und Didaktik, erst recht aber pädagogische Konzepte sind keine Techniken, die situationsübergreifend funktionieren. So setzen Lehrer\*innen pädagogische Programme ganz unterschiedlich um. Wer in verschiedene Klassen hineinschaut, die mit demselben Schulbuch, selbst mit demselben, in kleinen Schritten ausgearbeiteten Lehrgang arbeiten, wird einen ganz unterschiedlichen Unterricht erleben. Umgekehrt nutzen Lehrer\*innen auch spezifische Methoden im Rahmen ihrer jeweiligen didaktischen Konzeption auf verschiedene Weise: Freies Schreiben mit einer Anlauttabelle hat in REICHENS Konzeption „Lesen durch Schreiben“ eine ganz andere Bedeutung als im Vier-Säulen-Modell des Spracherfahrungsansatzes oder als Anhängsel eines Fibellehrgangs. Diese Konzept- und Kontextabhängigkeit der Umsetzung beeinflusst nicht nur die reale Intervention, sondern auch deren Wirkungen. Anders gesagt: Die Lehrperson

hat eine Schlüsselrolle bei der Übersetzung des Curriculums auf dem Papier in ein Curriculum in Aktion. Es stellt sich damit die Frage nach dem Alltagswert von Studien, die versuchen, den Einfluss der Randbedingungen möglichst weitgehend auszuklammern.

Für deren Anwendung auf den Einzelfall stellt sich ein zusätzliches Problem: Befunde aus Stichproben sind immer Wahrscheinlichkeitsaussagen. Folgt man ihnen, hat man die Sicherheit, in hundert oder gar tausend Fällen weniger Fehler zu machen, als wenn man sie nicht beachtet. Ob sie aber auf den Einzelfall zutreffen, kann man vorher nicht sagen. In den Mittelwerten der Stichprobenergebnisse werden die – in der Regel breit streuenden – Einzelwerte gemittelt. Die einzelne Lehrerin aber hat es mit einem konkreten Kind zu tun, und sie muss entscheiden, ob ihre Ayse oder ihr Linus ein im Sinne der Statistik „typisches“ Kind ist oder nicht. Als Vereinfachungen erleichtern solche Typisierungen das Alltags Handeln unter Zeitdruck – mit dem Risiko von Fehlentscheidungen im Einzelfall. Unter welchen Bedingungen eine Allgemeinaussage auf einen konkreten Schüler zutrifft – z. B. bei der Übergangsentscheidung am Ende der 4. Klasse – das sagen uns Leistungsstudien wie IGLU oder Testergebnisse aus den Vergleichsarbeiten (VerA) nicht.

Zu Recht wird kritisiert, wer glaubt, aus Einzelbeobachtungen generalisierbares Wissen ableiten zu können. Aber weniger beachtet wird der umgekehrte, mindestens gleich gewichtige Irrtum: Aus statistischem Wissen könne man Fallwissen für praktisches Handeln im Alltag direkt ableiten. Denn für Pädagogen besteht das eigentliche Problem darin, allgemeine Aussagen – seien es empirische Befunde der Forschung oder normative Vorgaben wie Lehrpläne – auf ihre jeweilige besondere Situation hin auszulegen.

Man stelle sich analog zum Vergleich von zwei Unterrichtsmethoden eine Untersuchung von Augenärzt\*innen bzw. Optiker\*innen vor, die die Sehfähigkeit von Grundschulkindern mit zwei Brillen vergleichen, eine mit Gläsern, die auf +2 Dioptrien geschliffen sind, und eine mit Minus-2-Gläsern. Da in dem Alter Kurzsichtigkeit häufiger ist, würde im Durchschnitt die Minus-2-Brille besser abschneiden – und zwar statistisch signifikant, wenn die Stichproben nicht zu klein sind. Darf man auf Grund dieses Befundes allen Kindern dieselbe Brille verordnen? Und selbst wenn die Studie nach Kurzsichtigen und Weitsichtigen unterschiede, würde doch niemand akzeptieren, dass aufgrund der unterschiedlichen Effekte ersteren eine Minus-2-Brille und letzteren eine Plus-2-Brille verschrieben würde. Denn wir wissen um die große Streuung der Fehlsichtigkeiten selbst innerhalb der Gruppen Kurz- und Weitsichtigkeit, da diese ja von 0,5 bis 6 und mehr variieren kann und da es zudem noch weitere Eigenschaften des Auges gibt, die über die Passung einer Brille entscheiden. Von einer Augenärzt\*in oder einem



Optiker\*in erwarten wir deshalb, dass sie die Gläser für die einzelnen Patient\*innen auswählen und gemeinsam mit ihnen ausprobieren, welche am besten passen. Denn über die bloße Dioptrienzahl hinaus sind weitere Aspekte – bis hin zum subjektiven Sehgefühl – zu berücksichtigen. In der Didaktik aber wird eine „evidenzbasierte Forschung“ als „Goldstandard“ akzeptiert, die aufgrund von Mittelwertvergleichen Methoden generell vorschreibt oder ausschließt, obwohl deren Wirkungen je nach Kontext breit streuen. Da ist der Erfahrungstransfer über Analogie („von Fall zu Fall“) nicht fehleranfälliger – dank seines Kontextreichtums oft sogar leichter auf Passung einzuschätzen als abstrakte Kennwerte statistischer Verrechnungen von Variablen.

#### **Zwischenfazit:**

Befunde in den Humanwissenschaften sind in hohem Maß situationsbedingt. Das Gewicht eines Faktors verändert sich von Fall zu Fall mit dem Einfluss anderer Faktoren.

Das heißt: Die Bedeutung eines menschlichen Verhaltens wechselt, sie ist kultur-, ja sogar Biografie abhängig. Diese Qualitäten werden in den Verallgemeinerungen der Großstudien neutralisiert – oder durch die Standardisierung der Erhebungen gar nicht erst erfasst. Die verallgemeinernde Absicherung von Befunden sichert somit keine Übertragbarkeit auf neue Situationen oder Einzelfälle. Nicht das arithmetische Mittel, sondern die Streuung der Einzelwerte um den Durchschnitt ist häufig das praxisrelevantere Datum. Lehrer\*innen können Befunde aus Großstudien zwar als Hypothesen nutzen, um ihre Aufmerksamkeit auf möglicherweise bedeutsame Bedingungen zu richten. Ob eher das Potenzial oder die Risiken einer Entscheidung zum Tragen kommen, muss in jedem Einzelfall neu überprüft werden. Insofern erweist sich die besondere Stärke statistischer Aussagen zugleich als ihre grundlegende Schwäche.

Entscheidungsexperimente funktionieren in der Didaktik nicht, weil sich die Vielfalt der Faktoren nicht kontrollieren lässt. Selbst wenn eine Methode im Durchschnitt um 5- oder 10-Prozentpunkte bessere Ergebnisse erreicht, ist die Streuung – schon durch die je nach Lehrer und Situation variierende Umsetzung – innerhalb jeder Methode so breit, dass die *within variance* immer größer ist als die *between variance*. Am Beispiel der Medizin veranschaulicht: Würde man allen die im Durchschnitt erfolgreichste OP-Methode vorschreiben, könnte das bedeuten, eine anspruchsvollere Methode den wenigen mit ihr erfolgreichen Operateur\*innen zu verbieten, weil die übrigen Ärzt\*innen mit traditionellen Verfahren besser zurecht kommen.

Insofern müssen wir im Umgang mit Forschungsbefunden umdenken: Statt zwischen "richtig" vs. "falsch" zu entscheiden, haben Wissenschaft und Praxis zu lernen, in den Kategorien "Potenziale" und "Risiken" zu denken - bezogen auf jeden Ansatz

#### **Die Objektivitätsillusion: Menschliches Verhalten ist mehrdeutig**

Auf den ersten Blick lassen die vielen empirischen Befunde zur Fehleranfälligkeit des Lehrerurteils bei der Bewertung von Schülerleistungen es attraktiv erscheinen, an seine Stelle standardisierte Tests zu setzen. Denn immer wieder haben Studien zur Vergabe von Noten bestätigt, was Alltagserfahrung von Schüler\*innen ist (vgl. Arbeitsgruppe Primarstufe 2014):

- Verschiedene Lehrer\*innen beurteilen dieselbe Arbeit – selbst in Mathematik – auf der üblichen 6er-Skala mit bis zu fünf unterschiedlichen Noten (mangelnde Objektivität).
- Eine Arbeit wird nach wenigen Monaten nur von einer Minderheit der Lehrer\*innen mit derselben Note bewertet wie beim ersten Mal (mangelnde Reliabilität).
- Noten weichen nicht zufällig von Testergebnissen ab, sondern sind je nach Sympathie oder Antipathie systematisch verzerrt (mangelnde Validität).
- Beim Wechsel in eine andere Klasse mit einem anderen Leistungsniveau kann sich die Bewertung derselben Schülerin um mehrere Notenstufen verbessern oder verschlechtern (mangelnde Vergleichbarkeit).

Kein Wunder also, dass man nach Alternativen sucht. Die Hoffnung heißt: standardisierte Tests. Sowohl die Erhebung als auch die Auswertung von Daten wird vorweg festgelegt („operationalisiert“), um sie von den beteiligten Personen unabhängig zu machen. Autor\*innen eines Leistungstests sollen Aufgaben demnach so gestalten, dass sie von allen getesteten Personen gleich verstanden und dass die Lösungen von verschiedenen Beurteilern gleich klassifiziert bzw. bewertet werden.

Die grundlegende Annahme des Standardisierungsparadigmas: Objektiv definiertes, beobachtbares Verhalten lässt sich eindeutig bestimmten Tiefenstrukturen (Kompetenzen bzw. Intentionen) zuordnen. Diese Annahme ist brüchig. Menschliches Verhalten ist immer mehrdeutig – bedingt durch den kulturellen und situativen Kontext und abhängig von Erwartungen und Deutungen der Beteiligten.

Dass jemand „die Hand hebt“, werden verschiedene Beobachter im Regelfall übereinstimmend feststellen. Wie sie dieses Verhalten deuten, das heißt, welche Absicht oder Aussage sie der beobachteten Person unterstellen, hängt aber vom Kontext ab:

- »Hier bin ich, komm her!« (auf dem Bahnsteig)
- »Ich stimme der Vorlage zu« (bei einem Wahlgang)
- »Bin bereit, zu dem angebotenen Preis zu kaufen« (bei einer Versteigerung)
- »Ich weiß die Antwort« (in einer Schulklasse).

Am Beispiel des Meldens im Unterricht lässt sich die Problematik erweitern. Auch wenn der Kontext klar ist, können Intention einerseits und Deutung andererseits immer noch auseinanderfallen. Manche Schüler\*innen melden sich, obwohl sie die Antwort nicht wissen – in der Hoffnung, dass die Lehrperson ihre Meldung im üblichen Sinne deutet und sie deshalb nicht aufruft. Es soll sogar Lehramtsanwärter\*innen geben, die ihre Klasse vor einer Lehrprobe instruieren, mit dem rechten Arm aufzuzeigen, wenn sie die Antwort wissen, und mit dem linken, wenn nicht. So können sie den uneingeweihten Prüfer\*innen eine „breite aktive Mitarbeit“ vortäuschen.

Nun lässt sich einwenden, dass diese Beispiele allenfalls für Befragungen relevant sind, etwa weil Personen dazu neigen, sozial erwünschte (genauer: von ihnen individuell als sozial erwünscht eingeschätzte) Antworten anzukreuzen. Aber bei der Bewertung von Leistungen stellt sich das Problem der mehrdeutigen Beziehung zwischen Oberflächen- und Tiefenstruktur analog. Insofern war es kühn von einer deutschen Kultusministerin, Eltern zu versprechen: „Jetzt wissen Sie wirklich, wo ihr Kind steht!“, als die Vergleichsarbeiten VerA in Klasse 3 flächendeckend eingeführt wurden. Denn die richtige Antwort zu einer Aufgabe kann aus verschiedenen Gründen gewählt worden sein, z. B.

- weil Schüler\*innen wie erwartet das von den Testautor\*innen avisierte Wissen und Können zur aktuellen Problemlösung genutzt haben, aber auch,
- weil sie diese oder eine ähnliche Aufgabe vor Kurzem bearbeitet und die verfügbare Lösung nur abgerufen oder analog entwickelt haben,
- weil sie dank ihrer allgemeinen Intelligenz durch eigenes Nachdenken die Lösung gefunden haben,
- weil sie unter Zeitdruck rasch noch ein Kreuz (zufällig richtig) gesetzt haben,
- weil sie die Lösung aus Hinweisen im Text und den verfügbaren Alternativen als plausibel erraten haben.

Umgekehrt gilt auch für Fehler, dass sie unterschiedliche Gründe haben können: Unaufmerksamkeit, unkonventioneller Lösungsweg usw. Vor allem aber gibt es bessere und schlechtere Fehler – und falsche Ergebnisse, die auf ein höheres Kompetenzniveau verweisen als eine richtige Lösung aus falschem Grund.

Ein Beispiel aus dem Anfangsunterricht, anhand einer vergleichsweise einfachen Aufgabe. Sechs Kinder haben geschrieben:

KINO	KINO	KINNO	KIHNOO	KIENO	KINO
------	------	-------	--------	-------	------

Bei der Auswertung eines standardisierten Tests würde herauskommen: Zwei Kinder haben das Wort falsch geschrieben und drei richtig – eine objektive und reliable, das heißt von beliebigen Testern reproduzierbare Aus- und Bewertung. Aber ist sie auch valide?

Unabhängig von Person und Situation ist nur eine Oberflächendeutung möglich. Um die Kompetenz der Kinder beurteilen zu können, muss man aber wissen, wie sie sonst schreiben und wie sie vorher geschrieben haben, das heißt auf welcher Stufe der Schreibentwicklung sie sich befinden. Aufgrund einer solchen – theoriegeleitet begleitenden – Lernbeobachtung könnten sich ganz andere Einschätzungen ergeben:

*Azur*, das erste Kind, schreibt KINO (richtig), weil er in einer Straße wohnt, in der er beim täglichen Gang zum Kindergarten an einem Kino vorbeikommt, auf dessen Schild ihn seine Mutter mehrfach hingewiesen hat. Auf ähnliche Weise hat er ein Dutzend Wörter gelernt, die er als Buchstabenfolge reproduzieren kann, ohne zu wissen, warum man z. B. KINO gerade mit diesen Buchstaben schreibt.

*Lea* schreibt mithilfe einer Anlauttabelle (HUNT, FARAT usw.). Sie hat das Wort nach ihrer Aussprache verschriftet – und Glück gehabt, dass diese Strategie im konkreten Fall zum richtigen Ergebnis führte. Sie schreibt also aus falschem Grunde richtig.

*Carl* ist mit KINNO schon einen Schritt weiter als Nora, weil er weiß, dass es nicht reicht, Laut für Laut zu verschriften. Man muss beispielsweise aufpassen, ob ein Selbstlaut lang oder kurz gesprochen wird, und dass dann „da etwas Besonderes ist“. Aber ob man einen Buchstaben verdoppelt – und ggf. welchen –, das hat er noch nicht verstanden. Carl schreibt also aus halbrichtigem Grund falsch. Oder er kommt aus einer Gegend, wo man den Vokal in VATTER und eben auch KINNO kurz spricht, dann wäre die Schreibung zwar falsch, aber orthographisch „legal“.

Legal sind auch die Rechtschreibmuster in KIHNOO. *Igor* hat nicht nur eine Reihe von orthographischen Besonderheiten wahrgenommen, sondern auch die Bedingungen erkannt, unter denen sie auftreten können. Er weiß aber noch nicht, dass im Deutschen der Langvokal im Regelfall nicht markiert wird – außer beim langen /i:/ durch das <ie>.

*Reha* dagegen hat schon gemerkt: Wenn man ein langes /i:/ hört und das Wort nicht kennt, dann ist es fast immer richtig und deshalb klug,



<ie> zu schreiben. Diese statistisch begründete Faustregel hilft ihr Fehler zu vermeiden. Sie schreibt also aus richtigem Grund falsch.

Leonie kennt diese Faustregel auch. Aber sie hat darüber hinaus gelernt, dass es Ausnahmen gibt wie MASCHINE, APFELSINE und eben KINO, die man sich als Einzelwörter merken muss. Nur sie schreibt aus richtigem Grund richtig.

Bei anspruchsvolleren Aufgaben nehmen die Interpretationsspielräume entsprechend zu. Kinder können schon die Aufgabe unterschiedlich verstehen – und die Auswertung kann leicht verkennen, was sich hinter der individuellen Lösung verbirgt. Es geht nicht um „falsch“ vs. „richtig“ auf der Leistungsebene, sondern um das Niveau der zugrunde liegenden Kompetenzen. Diese Probleme verschärfen sich im musisch-ästhetischen Bereich, bei sozialen und politischen Kompetenzen – generell: je höher die Anforderungen an das individuelle Denk- und Urteilsvermögen steigen, also je mehr es um Bildung und nicht nur um Leistung geht. Leider werden diese beiden Dimensionen nicht immer sauber getrennt – z. B. wenn die Kultusministerkonferenz ihre Kompetenzvorgaben „Bildungsstandards“ nennt, die eigentlich „Leistungsstandards“ heißen müssten, da sie sich nur auf fachliche Kompetenzen beziehen. Die Qualität grundlegender schriftsprachlicher und mathematischer Kompetenzen hängt zudem davon ab, wie, das heißt über welche konkreten Erfahrungen sie erworben wurden, in welche persönliche Gesamtkompetenz sie eingebunden sind.

Was hier für die Überprüfung des Lernerfolgs beschrieben wurde, spielt auch eine Rolle für das Lernen selbst. Sogar für eine scheinbar nur technische Leistung wie die Rechtschreibung lässt sich die Bedeutung der persönlichen Bedeutung (hier: von Wörtern) zeigen (vgl. die Beiträge zu Richter/Brügelmann 1994, Teil III). In entsprechenden Untersuchungen schneiden Jungen erwartungsgemäß im Durchschnitt schlechter ab als Mädchen. In der Regel nimmt der Rückstand der Jungen jedoch ab, wenn die Wörter aus ihrem Erfahrungs- und Interessensbereich stammen, z. B. bei „Schiedsrichter“, „Computer“ oder „Fahrrad“. Das letzte Wort ist besonders interessant, weil in einer Schweizer Teilstichprobe kein Aufholeffekt der Jungen festzustellen war. Diese scheinbare Irritation ist aber gerade ein Beleg für das angenommene Muster, ist in der Deutschschweiz doch das Wort „Velo“ üblich. Was hier als ein geschlechtsgruppenspezifischer Effekt sichtbar geworden ist, zeigt wie durch eine Lupe ein grundsätzliches Problem auf: die unterschiedliche Bedeutung derselben Aufgabe für das Lernen wie auch für dessen Überprüfung.

Diese Mehrdeutigkeit von Aufgaben und ihrer Lösung bezieht sich zunächst auf die unterste Ebene einer Begriffspyramide, die im Forschungsprozess dann immer abstrakter und damit noch interpreta-

tionsbedürftiger wird. Die unterschiedlichen Auslegungen „derselben“ Unterrichtsmethode durch verschiedene Lehrer\*innen habe ich bereits angesprochen. In Untersuchungen werden sie dennoch ein- und derselben Intervention zugeordnet. In einer Meta-Analyse wiederum werden verschiedene Studien zusammengefasst, die dieselbe Methode, z. B. „entdeckendes Lernen“, untersucht haben. Unberücksichtigt bleibt dabei, dass auch Forscher\*innen unterschiedliche Vorstellungen von diesem Konzept haben. Noch diffuser wird ein solcher Begriff in einer Meta-Meta-Analyse, wie HATTIE (2013) sie vorgelegt hat (siehe zur grundsätzlich sehr eingeschränkten Aussagekraft solcher (Meta-)Meta-Analysen Kap. 3.1 in Brügelmann 2015). So entstehen Sammelbegriffe wie „computerunterstützter Unterricht“, die in sich ähnlich heterogen sind wie auf der anderen Seite „Gruppenarbeit“ oder „offener Unterricht“ (s. ausführlicher zu der sehr eingeschränkten Aussagekraft solcher (Meta-)Meta-Analysen die Beiträge zu Terhart 2014). Damit verschärft sich das Problem der Interpretationsbedürftigkeit von Ergebnissen – und die Frage, wer die Deutungsmacht hat bzw. welche Möglichkeiten für eine unabhängige Urteilsbildung eröffnet werden.

Großstudien und standardisierte Erhebungen können für bestimmte Zwecke hilfreich sein – auch in der Pädagogik. Aber sie sind einer selbstkritischen Alltagserfahrung und einer sorgfältigen Beobachtung von Einzelfällen nicht grundsätzlich überlegen. Um den Nutzen empirischer Studien zu begründen, darf man zudem gute Forschung nicht mit schlechter Praxis vergleichen. In konkreten Handlungssituationen sind erfahrene Praktiker\*innen klugen Forscher\*innen meist überlegen. Weder ist Bildungsforschung zu Vorschriften berechtigt noch ist sie in der Lage, Handlungsentscheidungen aus Daten abzuleiten. Ihre Aufgabe ist Anregung, Herausforderung und kritische Befragung.

Angemessen sind weder eine Expertokratie noch eine bürokratische Forschungsorganisation; notwendig sind – im Sinne einer demokratiedienlichen Forschung – Transparenz, Machtverteilung und Dialog. Wo die technische Präzisierung von Methoden an ihre Grenzen stößt, sind Mechanismen sozialer Kontrolle erforderlich – sowohl bei der Unterrichts- und Schulentwicklung als auch bei der Leistungsrückmeldung auf Individualebene (vgl. unten die konkreten Vorschläge zu einer „Pädagogische Leistungskultur“). Dialogische Verfahren bieten keine Sicherheit für die Richtigkeit von Deutungen, aber Teilsichten können infrage gestellt, Autoritätsansprüche gebrochen werden. Ein Grundproblem der Sozialwissenschaften: Keine Methode kann gleichzeitig technische Genauigkeit/Objektivität und kulturelle Bedeutsamkeit/Gültigkeit von Beobachtungen sichern.

Diese Relativierung des Geltungsanspruchs von standardisierten Instrumenten ist keine Aufforderung, Ergebnisse von Tests oder von Befra-



gungen wegen der genannten Schwächen *wegzuwerfen*. Die Betonung ihrer Risiken sollte nicht dazu führen, die Potenziale standardisiert-statistischer Bildungsforschung aus dem Blick zu verlieren:

- Herausforderung und Kritik von (scheinbaren) Selbstverständlichkeiten der Alltagsroutinen durch den methodisch geschulten Fremdblick;
- Anregung und Unterstützung der Praxis durch systematisch geprüfte Hypothesen;
- Zufallskontrolle von Befunden durch repräsentative Stichproben;
- Reduktion komplexer Information auf handliche Kennwerte zur Groborientierung.

Tests können also hilfreiche Instrumente sein, wenn man sie als eine zusätzliche (nicht höhere) Sicht auf Schülerleistungen nutzt. Schon aus ökonomischen Gründen macht es Sinn, von Zeit zu Zeit die ganze Klasse mit Kurzverfahren zu testen, um bei Abweichungen von der eigenen Einschätzung genauer und differenzierter zu beobachten. Aber Lehrer\*innen und Eltern sollten sich ihren Ergebnissen auch nicht *unterwerfen*. Tests können reflektierte Erfahrung und Urteilskraft nicht ersetzen. Das hat auch Konsequenzen für die Lernbeobachtung und Leistungsbeurteilung im Unterricht.

#### **Pädagogische Lernkultur: Selbst- und Mitbestimmung im Unterricht**

„Offener Unterricht“ wird sowohl in der Praxis als auch bei seiner Evaluation durch Forscher/innen oft missverstanden als eine bloß *organisatorische* Flexibilisierung des Unterrichts. Gegenüber der herkömmlichen Differenzierung „von oben“ (durch die Lehrperson bzw. das Lehrwerk, z. Bv. durch einen Wochenplan mit vorgegebenen Aufgaben für drei Leistungsgruppen) ist es in der Tat ein Fortschritt, wenn Schüler/innen selbst wählen können, wie sie die anstehenden Aufgaben bearbeiten, z. B. die Reihenfolge, das Tempo, die Arbeits- und Sozialform. Aber die Kinderrechte fordern mehr (ausführlich Brügelmann 2005, Kap. 2 und Krappmann 2016, 36ff.). Anders als bei einer bloß organisatorischen Flexibilisierung verlangt eine *inhaltliche* Öffnung des Unterrichts, auch die Qualität der Aufgaben, Aktivitäten, Situationen zu verändern – mit dem Ziel, dass sie erlauben und verlangen,

- eigene Erfahrungen, Vermutungen, Strategien einzubringen/zu erproben und neue Erfahrungen in das eigene Denken einzuordnen;
- sich mit den Deutungen anderer Personen/Traditionen auseinander zu setzen, eine abweichende Sicht zu begründen;
- individuelle Vorstellungen/Zugangsweisen über verschiedene Konventionen bzw. fachspezifischen Theorien zu entwickeln.

Die Schweizer Didaktiker GALLIN/RUF (1999) haben für einen solchen Unterricht, der – auf allen Schulstufen – von den Erfahrungen und Vorstellungen der einzelnen Schüler/innen ausgeht, diese ins Gespräch miteinander bringt und erst daraus Konventionen und Routinen entwickelt, die einprägsame Formel „vom Singulären über das Divergierende zum Regulären“ gefunden. Dazu heißt es im Manifest „Kinderrechte, Demokratie und Schule“ (Krappmann 2016), (§ 55): *„Die Schulklasse und die Lerngruppe ermöglichen die Auseinandersetzung unter einer Vielzahl von Kindern, die fragen, bestreiten und entdecken, wie Fakten, Regeln, Zusammenhänge, Annahmen und Beweise zu verstehen sind, welche Folgerungen aus ihnen zu ziehen sind und welchen Sinn sie haben. Vor allem wird auch deutlich, dass es verschiedene Meinungen und Prioritäten gibt. Diese unterschiedliche Wahrnehmung, gelenkt von Interessen und Erfahrungen, entsteht in der Kommunikation und Kooperation mit anderen. Diese anderen sind zugleich Gegenüber und Partner für Gespräch, Widerspruch und Streit, aber auch für die Suche nach Klärung und Konsens“* (s. auch § 57, 59).

Noch einen Schritt weiter als diese inhaltliche Öffnung geht die ebenfalls im Manifest (Krappmann 2016, §§ 36, 398, 53) begründete Forderung, Schüler/innen auch an Entscheidungen über Ziele, Inhalte und Formen des Unterrichts zu beteiligen, z. B.

- individuell in Form von „Lernverträgen“ oder Wochenplänen, die lediglich Rahmenvorgaben für Aktivitäten machen, die die Kinder selbst bestimmen,
- gemeinsam über in der Gruppe entwickelte Projekte oder
- institutionell als Vereinbarung von sozialen Regeln im Klassenrat.

Kriterien für eine solche *pädagogische* Öffnung des Unterrichts lassen sich durch folgende Fragen näher bestimmen: Wie weit bietet der Unterricht den Schüler/innen ernsthafte Gelegenheiten,

- eigene Ziele und inhaltliche Schwerpunkte für ihre Arbeit zu setzen;
- an der Planung/ Gestaltung gemeinsamer Unterrichtsvorhaben mitzuwirken;
- Mitverantwortung für das soziale Zusammenleben in der Klasse und Schule zu übernehmen?

Konkreter formuliert – und durch Standards für den Unterrichtsalltag ergänzt – findet sich diese Position in den „Grundüberzeugungen“ des Schulverbunds der Reformschulen „Blick über den Zaun“ (Groeben u. a. 2005).

Diese Sicht auf Unterricht und Schule bestimmt auch die Kriterien des Deutschen Schulpreises (s. Fauser u. a. 2007) – eine wichtige Erweiterung des auf Fachleistungen eingeeengten Qualitätsblicks der KMK-Standards für die Hauptfächer. In unserem selektiven Schulsystem wer-

den diese zudem leicht zu Hürden, die zur Aussonderung von SchülerInnen führen. Darum hat der Grundschulverband (2005) seine Standards für guten Unterricht bewusst aus der Gegenperspektive als „Bildungsansprüche der Kinder“ formuliert.

Eine solche Lernkultur ist angewiesen auf Formen des Schullebens, wie sie im Entwurf angesprochen werden. Sie wird aber scheitern, wenn sich nicht auch die Formen der Leistungsbewertung ändern (vgl. Schönig 2008).

Spätestens seit dem Forschungsüberblick von INGENKAMP (1971) ist die begrenzte Aussagekraft von Ziffernnoten bekannt. Hinzu kommen pädagogisch unerwünschte Nebenwirkungen, wie sie in anschließenden Studien wiederholt belegt worden sind (vgl. die Zusammenfassung in: Arbeitsgruppe Primarstufe 2006). Der Versuch, diese Mängel durch den Einsatz von standardisierten Tests auszugleichen, scheitert an der Interpretationsbedürftigkeit von Aufgaben und Lösungen (s. o., ausführlicher zur Kritik dieses „Output“-Modells von Schule und Unterricht: Brügelmann 2015).

Zudem ist dieser Ansatz in der Regel mit einer Vorstellung von Diagnose und Förderung verknüpft, die ich vorne als Differenzierung „von oben“ beschrieben habe. Als Alternative zu dieser Sicht hat der Grundschulverband seine Konzeption „Pädagogische Leistungskultur“ und konkrete Materialien für eine dialogische Form der Wahrnehmung und Einschätzung von Leistungen in allen Fächern und über die gesamte Grundschulzeitschrift hinweg entwickelt (vgl. Bartnitzky u. a. 2005ff.). Ihre Leitideen:

Leistungsanforderungen sind gemeinsam zu klären und in *persönliche Zielvereinbarungen* zu übersetzen, d. h. mit jedem Kind werden für seinen Entwicklungsstand anspruchsvolle Anforderungen verabredet. Für die Schulen, die es mit der Inklusion ernst meinen, bedeutet das: gleiche Zielperspektiven für alle, aber jeweils unterschiedlich hoch angesetzte Zwischenziele. Dies setzt als erstes voraus, dass Ziele und Inhalte der Lehrpläne in eine verständliche Sprache und Form übersetzt werden (z. B. in Form einer Lernlandkarte).

Um jeweils den passenden „nächsten Schritt“ zu ermöglichen, müssen zudem gehaltvolle, *offene Aufgaben* gestellt werden, die unterschiedliche Anforderungen stellen und damit individuelle Lernwege zulassen. In der inklusiven Schule ist das besonders deutlich: Ihre Pädagogik besteht nicht darin, einige Kinder mit besonderen Handicaps in die Lerngruppe aufzunehmen, die dann spezielle Aufgaben erhalten. Es geht vielmehr um eine andere Sicht auf *alle* Kinder. Jedes Kind ist besonders und braucht den Raum, sein Können zu zeigen und es weiterzuentwickeln. Nicht die Aufgabe variiert,

sondern die Fehlertoleranz, der erforderliche Zeitbedarf und die gewährte Unterstützung. Auf dem jeweiligen orthographischen Niveau geschriebene eigene Texte, die durch eine Übersetzung in „Buchschrift“ ergänzt werden, erlauben schon in der ersten Klasse Kindern mit sehr unterschiedlichen Voraussetzungen, ihre Vorstellungen und Erfahrungen zu Papier zu bringen und sie anderen mitzuteilen. Wer noch gar nicht lesbar schreiben kann, diktiert der Lehrperson oder einem schriftkundigen Kind seine Geschichte. Auch im Mathematikunterricht sind offene Aufgaben denkbar, die ganz unterschiedliche Zugänge ermöglichen (wie z. B. die sog. „Fermi-Aufgaben“, s. für die Grundschule auch die Beispiele aus dem Dortmunder Projekt „PIKAS“ von Selter u. a. im Internet).

Basis für die damit erforderliche *individuelle Lerndokumentation* (z.B. Lese-Portfolios) sind begleitende Beobachtungen und Gespräche mit den Kindern über ihre Fortschritte und über spezifische Schwierigkeiten bei der Auseinandersetzung mit den Lerninhalten. Statt gemeinsamer Klassenarbeiten stellen sich die Kinder bestimmten Prüfungsanforderungen, wenn sie sich den Anforderungen gewachsen fühlen – wie beim „Seepferdchen“ oder später in der Fahrschule. Die nachgewiesenen Kompetenzen können in Ausweisen (z. B. in Rechenpässen mit gestuften Anforderungen, in einem Labor- oder Computer-Führerschein) dokumentiert werden.

Der Austausch von Selbst- und Fremdeinschätzungen ermöglicht eine *dialogische Leistungsbewertung*, z. B. über regelmäßige Gespräche der Kinder untereinander, zwischen dem Kind und der Lehrperson und auch – z. B. als Kombination von Zeugnisübergabe und Elternsprechtag – zwischen Kind, Lehrer/in und Eltern.

Mehrperspektivität und Dialog sind Grundprinzipien der Leistungsbewertung – und zwar von der ersten Klasse an und für alle Kinder (z. B. durch eine Kontrastierung von Selbst- und Fremdeinschätzung).

Diese Prinzipien bestimmen darüber hinaus die Lernkultur in der demokratischen Schule generell. Diese kann nicht mehr aus den didaktischen Traditionen der Bekehrung und der Belehrung leben. Sie muss sich verstehen als Ort der Begegnung – des Austausches zwischen den Generationen und den Kulturen (vgl. Krappmann 2016, § 68). Schule soll Fenster öffnen, um Kindern Zukünfte zu ermöglichen, die nicht durch den Zufall der Geburt beschränkt sind – aber auch nicht durch ein Standardbild noch so gut gemeinter Pädagogik.

## Literatur

Arbeitsgruppe Primarstufe (2006): Sind Noten nützlich und nötig? Zifferzensuren und ihre Alternativen im empirischen Vergleich. Eine wissenschaftliche Expertise des Grundschul-



verbandes, erstellt von der Arbeitsgruppe Primarstufe an der Universität Siegen. Grundschulverband e.V.: Frankfurt (aktualisiert 2014)

Backhaus, A./Knorre, S., in Zusammenarbeit mit Brügelmann, H., und Schiemann, E. (Hrsg.) (2008): Demokratische Grundschule - Mitbestimmung von Kindern über ihr Leben und Lernen. Arbeitsgruppe Primarstufe/ FB2. Verlag Universität: Siegen. Download: [http://www.pedocs.de/frontdoor.php?source\\_opus=13014](http://www.pedocs.de/frontdoor.php?source_opus=13014)

Bartnitzky, H., u. a. (Hrsg.) (2005&2006&2007): Pädagogische Leistungskultur: Materialien für Klasse 1/2 und Klasse 3/4. Beiträge zur Reform der Grundschule, Bd. 119 & 121 & 124. Grundschulverband: Frankfurt

Baumert, J. (2016): Leistungen, Leistungsfähigkeit und Leistungsgrenzen der empirischen Bildungsforschung. Das Beispiel von Large-Scale-Assessment-Studien zwischen Wissenschaft und Politik. In: Zeitschrift für Erziehungswissenschaft, Supplement I, Vol. 19, 215-253

Beutel, I./Beutel, W. (Hrsg.) (2008): Beteiligt oder bewertet? Leistungsbeurteilung und Demokratiepädagogik. Politik und Bildung, Bd. 58. Wochenschau Verlag: Schwalbach

Böttcher, W., u. a. (Hrsg.) (2009): Evidenzbasierte Bildung: Wirkungsevaluation in Bildungspolitik und pädagogischer Praxis. Waxmann: Münster

Brinkmann, E. (Hrsg.) (2015): Rechtschreiben in der Diskussion – Schriftspracherwerb und Rechtschreibunterricht. Beiträge zur Reform der Grundschule, Bd. 140, Grundschulverband: Frankfurt

Brinkmann, E. (2018): Der Spracherfahrungsansatz im Anfangsunterricht. In: Grundschule, 50. Jg., H. 6, 28-32

Brügelmann, H. (1972): Lernziele im offenen Curriculum. In: Thema Curriculum, 1. Jg., Nr. 2, 16-45

Derselbe (1999): From invention to convention. Children's different routes to literacy. How to teach reading and writing by construction vs. instruction. In: Nunes, T. (ed.) (1999): Learning to read: An integrated view from research and practice. Kluwer: Dordrecht et al., 315-342

Derselbe (2005): Schule verstehen und gestalten – Perspektiven der Forschung auf Probleme von Erziehung und Unterricht. Libelle: CH-Lengwil (bis 2008 aktualisiert unter: [www.agprim.uni-siegen.de/schuleverstehen](http://www.agprim.uni-siegen.de/schuleverstehen))

Derselbe (2015): Vermessene Pädagogik – standardisierte Schüler? Risiken und Nebenwirkungen einer „evidenzbasierten“ Bildungspolitik und Schulpraxis (Arbeitstitel). Beltz: Weinheim/ Basel)

Brügelmann, H./ Brinkmann, E. (2013): Freies Schreiben im Anfangsunterricht? Eine kritische Übersicht über Befunde der Forschung [1. Fassung 2006, aktualisiert und ergänzt November 2013]. Download: [www.grundschulverband.de/fileadmin/grundschulverband/Download/Forschung/Zur Leseforschung\\_Feb\\_07.pdf](http://www.grundschulverband.de/fileadmin/grundschulverband/Download/Forschung/Zur_Leseforschung_Feb_07.pdf)

Brügelmann, H./ Söhnen, R. (1981): Ist Lesen Lehren lernbar? Plädoyer für eine vergleichende Erprobung von Lehrgängen für den Anfangsunterricht. In: IRA/D Beiträge, 4. Jg., 1/1981, 14-18

Brügelmann, H., u. a. (Hrsg.) (1995): Am Rande der Schrift. Zwischen Mehrsprachigkeit und Analphabetismus. DGLS-Jahrbuch Bd. 6. Libelle Verlag: CH-Lengwil. u. a.

Deutscher Bildungsrat (1974): Zur Förderung praxisnaher Curriculumentwicklung. Empfehlungen der Bildungskommission. Bundesdruckerei: Bonn

Fausser, P., u. a. (Hrsg.) (2007): Was für Schulen! Portraits der Preisträgerschulen und der nominierten Schulen des Wettbewerbs 2006. Klett Kallmeyer: Stuttgart

Funke, R. (2014): Erstunterricht nach der Methode Lesen durch Schreiben und Ergebnisse schriftsprachlichen Lernens – Eine metaanalytische Bestandsaufnahme. In: Didaktik Deutsch, 19. Jg., H. 36, 20-41

Gallin, P./Ruf, U. (1990): Sprache und Mathematik in der Schule. Auf eigenen Wegen zur Fachkompetenz. Illustriert mit sechzehn Szenen aus der Biographie von Lernenden. Verlag Lehrerinnen und Lehrer Schweiz: Zürich (Neuaufgabe Friedrich/ Kallmeyer: Seelze 1998)

Groebe, A. v. d., u. a. (2005): Unsere Standards. Ein Diskussionsentwurf, vorgelegt von „Blick über den Zaun“ – Bündnis reformpädagogisch engagierter Schulen. In: Neue Sammlung, 45. Jg., H. 2, 253-297

Grotlüschen, A./Riekman, W. (Hrsg.) (2012): Funktionaler Analphabetismus in Deutschland. Ergebnisse der ersten leo. – Level-One Studie. Waxmann: Münster

Grundschulverband (2005): Bildungsansprüche von Grundschulkindern – Standards zeitgemäßer Grundschularbeit. In: Bartnitzky u. a. (2005, H. 1; Erstveröffentlichung als Heft Nr. 81 von „Grundschule aktuell“ 2003)

Hattie, J.A.C. (2013): Lernen sichtbar machen. Schneider Hohengehren: Baltmannsweiler (engl. 2009)

Hüttis-Graff, P. (2015): „Feuer mit i?“ – Wege zum Rechtschreiben in Klasse 1. In: Grundschulzeitschrift, 29. Jg., H. 290, 48-53

Hüttis-Graff, P./Wirsing, D. (2018): Anfänge des Rechtschreiblernens 1994 und 2014 – Ergebnisse einer diachronen Longitudinalstudie in Hamburg. In: Didaktik Deutsch, H. 45, 50-73

idw (2018): Die „Fibel“ führt zu besserer Rechtschreibung. Pressemitteilung der Rheinischen Friedrich-Wilhelms-Universität Bonn v. 10.09.2018

Ingenkamp, K. (Hrsg.) (1971): Die Fragwürdigkeit der Zensurengebung. Beltz: Weinheim/ Basel (9. Aufl. 1995)

Kirschhock, E.-M. (2004): Entwicklung schriftsprachlicher Kompetenzen im Anfangsunterricht. Klinkhardt: Bad Heilbrunn

Krappmann, L. (2016): Kinderrechte, Demokratie und Schule. Ein Manifest. In: Krappmann/ Petry (2016, 17-53)

Krappmann, L./Petry, C. (Hrsg.) (2016): Worauf Kinder und Jugendliche ein Recht haben. Kinderrechte, Demokratie und Schule: Ein Manifest. Debus-Verlag: Schwalbach

Kuhl, T./Röhr-Sendmeier, U. N. (2018): Rechtschreiberfolg nach unterschiedlichen Didaktiken. Eine kombinierte Längsschnitt-Querschnittstudie. Poster für den Bundeskongress der Schulpsychologie am 20.-22.9.2018 in Frankfurt

May, P. (1995): Schriftsprachliche Leistungen und lernförderliche Unterrichtsbedingungen. Ergebnisse der Voruntersuchung zum Hamburger Projekt "Lesen und Schreiben für alle". In: Brügelmann u. a. (1995, 344-349)

Maß, P. (2013): HSP 1 – 10. Hamburger Schreib-Probe. Manual/Handbuch: Diagnose orthografischer Kompetenz zur Erfassung der grundlegenden Rechtschreibkompetenzen. Neustandardisierung 2012. Verlag für Pädagogische Medien/Klett: Stuttgart

National Early Literacy Panel (Hrsg.) (2008): Developing early literacy: A scientific synthesis of early literacy development and implications for intervention. National Institute for Literacy & The Partnership for Reading. Jessup, Maryland

National Reading Panel (2000): Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. NICHD: Washington, D.C.

Peschel, F. (2002a+b): Offener Unterricht – Idee – Realität – Perspektive und ein praxiserprobtes Konzept zur Diskussion

Plewnia, A./Witt, A. (Hrsg.) (2014): Sprachverfall? Dynamik – Wandel – Variation. IDS Jahrbuch 2013. de Gruyter: Berlin u.a.

Ramseger, J. (1977): Offener Unterricht in der Erprobung. Erfahrungen mit einem didaktischen Modell. Juventa: München (3. Aufl. 1992)

Richter, S. (1992): Die Rechtschreibentwicklung im Anfangsunterricht und Möglichkeiten der Vorhersage ihrer Störungen. Phil. Diss. FB 12 der Universität Bremen. Verlag Dr. Kovac: Hamburg

Richter, S./ Brügemann, H. (Hrsg.) (1994): Mädchen lernen ANDERS lernen Jungen. Geschlechtsspezifische Unterschiede beim Schriftspracherwerb. DGLS-Reihe "Lesen und Schreiben". Libelle: CH-Lengwil (2. Aufl. 1996)

Ryan, R. M./Deci, E. L. (2000): Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. In: American Psychologist, Vol. 55, 68-78

Scheerer-Neumann, G./ Schnitzler, C. D. (2009): Rechtschreiberwerb im zweiten Schuljahr. In: Zeitschrift für Grundschulforschung, 2. Jg., H. 2, 95-110

Schönig, W. (2008): Demokratisierung der Schule durch eine pädagogisch akzentuierte Leistungsbeurteilung? Überlegungen zur Arbeit am Leistungsethos der Schule. In: Beutel/ Beutel (2008, 166-182)

Schründer-Lenzen, A./ Merkens, H. (2006): Differenzen schriftsprachlicher Kompetenzentwicklung bei Kindern mit und ohne Migrationshintergrund. In: Schründer-Lenzen (2006, 15-44)

Schründer-Lenzen, A. (Hrsg.) (2006): Risikofaktoren kindlicher Entwicklung. Migration, Leistungsangst und Schulübergang. VS Verlag Sozialwissenschaften: Wiesbaden 2006

Stanat, P., u. a. (2017): IQB-Bildungstrend 2016. Kompetenzen in den Fächern Deutsch und Mathematik am Ende der 4. Jahrgangsstufe im zweiten Ländervergleich. Waxmann: Münster/ New York

Steinig, W., u. a. (2009): Schreiben von Kindern im diachronen Vergleich. Texte von Viertklässlern aus den Jahren 1972 und 2002. Waxmann: Münster u. a.

Steinig, W./Betzel, D. (2014.): Schreiben Grundschüler heute wirklich schlechter als vor 40 Jahren? Texte von Viertklässlern aus den Jahren 1972, 2002 und 2012. In: Plewnia/ Witt (2014, 353-371)

Terhart, E. (Hrsg.) (2014): Die Hattie-Studie in der Diskussion. Probleme sichtbar machen. Klett Kallmeyer: Seelze

Valtin, R./Voss, A. (2010): Rechtschreiben und Förderung – aktuelle Ergebnisse aus IGLU In: Grundschule Deutsch, 7. Jg., H. 27, 42-43

## Anhang:

### Der Spracherfahrungsansatz: Freies Schreiben von Anfang an<sup>1</sup>

Erika Brinkmann

**Lesen- und Schreibenlernen ist ein individueller Prozess. Der Spracherfahrungsansatz nimmt die unterschiedlichen Vorerfahrungen der Kinder ernst und motiviert sie durch persönlich bedeutsame Begeg-**

<sup>1</sup> Erschienen in: Grundschule 50.Jg. (2018). Abdruck mit freundlicher Genehmigung der Autorin.